Bayesian Analysis (2024) **19**, Number 1, pp. 181–203

Decoupling Shrinkage and Selection in Gaussian Linear Factor Analysis[*]

Henrique Bolfarine[†], Carlos M. Carvalho[‡], Hedibert F. Lopes[§,¶],
and Jared S. Murray[‖]

Abstract. Factor analysis is a popular method for modeling dependence in multivariate data. However, determining the number of factors and obtaining a sparse orientation of the loadings are still major challenges. In this paper, we propose a decision-theoretic approach that brings to light the relationship between model fit, factor dimension, and sparse loadings. This relation is done through a summary of the information contained in the multivariate posterior. A two-step strategy is used in our method. First, given the posterior samples from the Bayesian factor analysis model, a series of point estimates with a decreasing number of factors and different levels of sparsity are recovered by minimizing an expected penalized loss function. Second, the degradation in model fit between the posterior of the full model and the recovered estimates is displayed in a summary. In this step, a criterion is proposed for selecting the factor model with the best trade-off between fit, sparseness, and factor dimension. The findings are illustrated through a simulation study and an application to personality data. We used different prior choices to show the flexibility of the proposed method.

Keywords: Bayesian factor analysis, model selection, sparse loadings, factor dimension, loss function.

1 Introduction

Factor analysis is an important tool for modeling the dependence structure among variables. Over the years, factor analysis and related factor models have found their way into applications in different fields, such as economics, finance, and genomics (see Fruehwirth-Schnatter and Lopes, 2018, and references therein). However, selecting the number of factors and generating a sparse and interpretable loading matrix can both be challenging tasks (Ročková and George, 2016). In this paper, we address these challenges by introducing a unique decision-theoretic approach (Bernardo and Smith, 2009)

[*] Henrique Bolfarine acknowledges the support from CAPES (Coordenação de Aperfeiçomento de Pessoal de Nível Superior), grant number 88887.571312/2020-00 and from the Salem Center for Policy at the University of Texas at Austin McCombs School of Business. Hedibert F. Lopes thank FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo) for financial support through grant number 2018/04156-9.

[†] McCombs Schools of Business, University of Texas at Austin, henrique.bolfarine@austin.utexas.edu

[‡] McCombs Schools of Business, University of Texas at Austin

[§] School of Mathematical and Statistical Sciences, Arizona State University

[¶] Also affiliated to Insper Institute of Education and Research, São Paulo, Brazil

[‖] McCombs Schools of Business, University of Texas at Austin

https://doi.org/10.1214/22-BA1349

that brings to light the relation between model fit, factor dimension, and a sparse representation of the loading matrix.

Our approach has two key objectives: to summarize the information in the multivariate posterior and to obtain interpretable point estimates for factor loadings and uniqueness from the decision analysis. To achieve this, we followed the procedure described by Hahn and Carvalho (2015). This method, known as decoupling shrinkage and selection (DSS), is a model selection strategy based on the posterior predictive distribution, which provides a meaningful scale on which to determine whether a sparse, lower-dimensional version of the model has a sufficient fit. The DSS method, which was originally developed for linear regression, has been successfully applied to a variety of statistical models, including seemingly unrelated regressions (Puelz et al., 2017), graphical models (Bashir et al., 2019), functional regressions (Kowal and Bourgeois, 2020), nonlinear regressions (Woody et al., 2021), time-varying parameter models (Huber et al., 2021), nonparametric item response theory (Krantsevich et al., 2021), and Bayesian additive regression trees (Carvalho et al., 2021).

We summarize the DSS for factor analysis (DSSFA) in two steps. First, provided that posterior samples from the Bayesian factor analysis model are available, a series of optimal point estimates with a decreasing number of factor dimensions and with different levels of sparsity in the loadings are obtained through the minimization of a penalized loss function. Second, we generate a posterior summary that encapsulates the loss in fit between the full factor model, produced by the posterior distribution, and the model generated by sparse lower dimension estimates. This summary is displayed in a plot that can be visually inspected in search of the model that yields the best fit. Nonetheless, we also propose a criterion that automatically selects the factor model with the best trade-off between fit, sparseness, and factor dimension.

The DSSFA approach connects different strands of the factor analysis literature. It incorporates ideas from parametric methods, where posterior samples are obtained via well-established stochastic algorithms (Lopes and West, 2004; West, 2003; Carvalho et al., 2008; Fruehwirth-Schnatter and Lopes, 2018), from methods that do not impose identifying restrictions on their inference algorithms, and do not require pre-specification of the factor dimension (Bhattacharya and Dunson, 2011; Legramanti et al., 2020), and from methods where sparse loadings play an important role in the estimation of the covariance matrix, resulting in parsimonious models (Nakajima and West, 2013; Kastner, 2019). In addition, unlike hard thresholding rules and classical information-based approaches (Schwarz, 1978; Akaike, 1987), our method allows for simultaneous selection of factor models with varied dimensions and loading matrices with different levels of sparsity. To our knowledge, this is the first study, to treat model selection for factor analysis as a decision problem and to apply ideas from the original DSS approach to latent variable modeling. In addition, we show a substantial runtime improvement over conventional Bayesian approaches without significant loss in fit.

This paper is organized as follows: The remainder of this section reviews the factor analysis model. In Section 2, we introduce the framework of the DSSFA method. In Section 3, we use the simulation design from Man and Culpepper (2022) to compare our approach to the marginal likelihood estimate for selecting the number of factors (Lopes and West, 2004; Newton and Raftery, 1994); and Bayes model averaging (BMA, Hoeting et al., 1999) for covariance matrix estimation. In Section 4, we apply our method to a subset of the big five personality traits data and assess how our approach interacts with over-fitted priors (Bhattacharya and Dunson, 2011; Legramanti et al., 2020), resulting in meaningful posterior summaries and interpretable factor loadings. Finally, some conclusions are given in Section 5. The methods and data presented here are available at `https://github.com/hbolfarine/dssfa`.

1.1 Notations for the basic factor analysis model

In the basic factor analysis model, $\boldsymbol{y}_i = (\boldsymbol{y}_{1i}, \ldots, \boldsymbol{y}_{pi})^T$ is a p-dimensional vector of observations in a random sample $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^T$, that relates to a k-dimensional vector of common latent factors $\boldsymbol{f}_i$, with $k \leq p$, through

$$\boldsymbol{y}_i = \boldsymbol{B}\boldsymbol{f}_i + \boldsymbol{\epsilon}_i, \tag{1.1}$$

where $\boldsymbol{B}$ is a $p \times k$ factor loadings matrix, $\boldsymbol{f}_i \sim N_k(\boldsymbol{0}, \boldsymbol{I}_k)$, and $\boldsymbol{\epsilon}_i$ is the idiosyncratic error vector with dimension p. We assume in model (1.1) that (i) $\boldsymbol{\epsilon}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ and $\sigma_j^2 > 0$, for all $j = 1, \ldots, p$, and (ii) $\boldsymbol{f}_r$ and $\boldsymbol{\epsilon}_t$ are independent for all r and t. These assumptions imply that the distribution of $\boldsymbol{y}_i$ conditioned to $\boldsymbol{f}_i$ is given by $N_p(\boldsymbol{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \boldsymbol{B}\boldsymbol{B}^T + \boldsymbol{\Sigma}$ is the covariance matrix (Thurstone, 1947).

To complete the Bayesian specification of model (1.1), we assume the prior distributions for the loadings and uniqueness are independent, $p(\boldsymbol{B}, \boldsymbol{\Sigma}) = p(\boldsymbol{B})p(\boldsymbol{\Sigma})$. A typical choice for $p(\boldsymbol{B})$ is the to use truncated normal priors for the diagonal components of the loading matrix and normal priors for the remaining lower triangular values. This setup is refereed to as positive lower triangular constraint (PLT, Geweke and Zhou, 1996; Lopes and West, 2004). Additionally, inverse-gamma priors are used for the uniqueness. Both of these choices are useful since they provide conjugate forms that are easy to compute using the Gibbs sampler (Gamerman and Lopes, 2006).

The goal of the PLT constrain is to address the problem of posterior identifiability of $\boldsymbol{B}$, in which one can obtain the same covariance matrix $\boldsymbol{\Omega}$, defined by (1.1), by multiplying $\boldsymbol{B}$ by an orthonormal matrix $\boldsymbol{P}$ where $\boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{I}_k$ (see Geweke and Zhou, 1996). An analysis of recent approaches to prior choices on the problem of posterior identifiability can be seen in Man and Culpepper (2022) and Papastamoulis and Ntzoufras (2022).

One of the most important aspects of our method, as will be discussed further in this study, is that no prior modeling assumptions are made as long as samples from the posterior marginal distributions, $p(\boldsymbol{B}|\boldsymbol{y})$, and $p(\boldsymbol{\Sigma}|\boldsymbol{y})$ are available. Our technique simply requires knowledge of the factor dimension of the posterior loadings, k. Throughout the rest of this paper, we will refer to the full model posterior, or full factor model, as the posterior on $\boldsymbol{B}$ with $p \times k$ dimensions.

2 Decoupling shrinkage and selection in factor analysis

In this section, we present the essential features of the DSSFA approach. First, we present how the sparse factor analysis estimates are obtained using the proposed decision framework. Second, we present the posterior summary, which exposes the loss in fit between the full factor model and the sparse estimates, as well as a model selection criterion. This section concludes with an overview of our technique and its application to a toy example.

2.1 DSSFA estimates

As it is well known, one of the main challenges in decision analysis is to select a proper loss function (Bernardo and Smith, 2009; Berger, 2013). In our famework, we opted for the negative loglikelihood of the multivariate normal distribution since this function depends uniquely on the covariance matrix and thus relates directly to model (1.1). Furthermore, Stein's loss and the Kullback-Leibler divergence between two normal distributions are also strongly associated with this loss (Dey and Srinivasan, 1985; Kullback, 1997). Other use of a similar loss function may be seen in Bashir et al. (2019), where it was applied to recover sparse estimates from a Gaussian graphical models.

To highlight the trade-off between fit and model simplicity, we add a complexity penalty, $P(\cdot)$, to the proposed loss, yielding

$$\mathcal{L}_\lambda(\boldsymbol{\Omega}, \tilde{\boldsymbol{\Omega}}) = \log|\tilde{\boldsymbol{\Omega}}| + \mathrm{tr}\left(\tilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega}\right) + \lambda P(\tilde{\boldsymbol{\Omega}}), \tag{2.1}$$

where $\tilde{\boldsymbol{\Omega}}$ is a $p \times p$ positive definite symmetric matrix, $\boldsymbol{\Omega}$ is a $p \times p$ covariance matrix defined by the assumptions presented in Section 1.1, $\mathrm{tr}(A)$ is the trace, and $|A|$ is the determinant of matrix A. In this setup, the complexity parameter $\lambda > 0$ controls the trade-off between model fit and parsimony (Hahn and Carvalho, 2015).

It is important to reiterate the distinction between $\tilde{\boldsymbol{\Omega}}$ and the covariance matrix $\boldsymbol{\Omega}$ in (2.1). As seen in Section 1.1, we have that both $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, and consequently $\boldsymbol{\Omega}$ are parameters of the Bayesian factor analysis model and thus are associated with the prior distribution. By comparison, it is not coherent to place a prior on $\tilde{\boldsymbol{\Omega}}$ and λ as they define an action and a penalty, respectively, in the proposed framework.

Usually, in the DSS framework, the posterior predictive distribution is used to obtain the best accuracy in prediction given future observations (Hahn and Carvalho, 2015; Woody et al., 2021; Kowal, 2021). In this paper, however, we focus largely on parameter rather than the predictive distribution since we were interested in model fit rather than predictions. Thus, by minimizing the posterior expectation of (2.1), over the posterior distribution of the factor analysis model results in

$$\hat{\boldsymbol{\Omega}}_\lambda \equiv \underset{\tilde{\boldsymbol{\Omega}}}{\mathrm{argmin}}\, E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[\mathcal{L}_\lambda(\boldsymbol{\Omega}, \tilde{\boldsymbol{\Omega}})\right], \tag{2.2}$$

where $\hat{\boldsymbol{\Omega}}_\lambda$ is the optimal point estimate, and $\boldsymbol{\theta} = (\boldsymbol{B}, \boldsymbol{\Sigma})$.

We simplify the expectation in expression (2.2), as follows

$$E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[\mathcal{L}_\lambda(\boldsymbol{\Omega}, \tilde{\boldsymbol{\Omega}})\right] \quad = \quad E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[\log|\tilde{\boldsymbol{\Omega}}| + \mathrm{tr}\left(\tilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega}\right)\right] + \lambda P(\tilde{\boldsymbol{\Omega}})$$

$$= \quad \log|\tilde{\boldsymbol{\Omega}}| + \operatorname{tr}\left(\tilde{\boldsymbol{\Omega}}^{-1}\overline{\boldsymbol{\Omega}}\right) + \lambda P(\tilde{\boldsymbol{\Omega}}), \tag{2.3}$$

where $\overline{\boldsymbol{\Omega}} = E_{\boldsymbol{\theta}|\boldsymbol{y}}[\boldsymbol{\Omega}]$ is the posterior mean of the covariance matrix. By integrating over the marginal posterior distributions of the Bayesian factor analysis model, $p(\boldsymbol{B}|\boldsymbol{y})$, and $p(\boldsymbol{\Sigma}|\boldsymbol{y})$, we have $\overline{\boldsymbol{\Omega}} = \overline{\boldsymbol{B}\boldsymbol{B}^T} + \overline{\boldsymbol{\Sigma}}$, where $\overline{\boldsymbol{B}\boldsymbol{B}^T} = E_{\boldsymbol{B}|\boldsymbol{y}}\left[\boldsymbol{B}\boldsymbol{B}^T\right]$, and $\overline{\boldsymbol{\Sigma}} = E_{\boldsymbol{\Sigma}|\boldsymbol{y}}[\boldsymbol{\Sigma}]$ are the expected posterior values.

We make some observations on equation (2.3): (i) The expected loss function depends uniquely on the posterior mean of the covariance matrix $\overline{\boldsymbol{\Omega}}$, so it can be applied in conjunction with different prior choices, (ii) $\overline{\boldsymbol{\Omega}}$ is robust to factor rotation, so the factor analysis model's identifiability on prior choice is not an immediate concern.

Since our interest lies in optimal decisions for the factor analysis model, with a special interest in the structure of the loadings matrix $\boldsymbol{B}$, we assume $\tilde{\boldsymbol{\Omega}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{B}}^T + \tilde{\boldsymbol{\Sigma}}$, with $\tilde{\boldsymbol{B}}$ is a $p \times \tilde{k}$ matrix, $\tilde{\boldsymbol{\Sigma}}$ is a $p \times p$ diagonal matrix, with positive entries, and $\tilde{k} \in \{1, 2, \dots\}$, which are all elements of the decision analysis. Notably, $\tilde{k}$ is a choice of the dimension on the resulting loadings estimates. For instance, for smaller values of $\tilde{k}$, the resulting optimal estimates have lower dimensions, and when $\tilde{k} = k$, results in an estimate with the dimension of the full model posterior. In this paper, no further restrictions are imposed on the identifiability of $\tilde{\boldsymbol{B}}$, although a $\tilde{k} \times \tilde{k}$ symmetric matrix $\tilde{\boldsymbol{\Phi}}$ could be introduced to the decomposition of $\tilde{\boldsymbol{\Omega}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{B}}^T + \tilde{\boldsymbol{\Sigma}}$, resulting in a decision analysis for the oblique factor model (Thurstone, 1947).

We reiterate that $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{k}$ are actions in the decision analysis, and thus are not subject to the prior specification of the Bayesian factor analysis model. By comparison, a prior on $\boldsymbol{B}$ may indicate our preference for sparse loadings, although it does not guarantee sparsity in the posterior (West, 2003; Carvalho et al., 2008). In contrast, we can use our framework to extend the complexity penalty $P(\cdot)$ to include $\tilde{\boldsymbol{B}}$, allowing for a sparse representation of the estimates regardless of prior choice. Many choices of complexity penalty are available, but to shrink the factor loadings to zero, we consider sparsity-inducing penalties such as ℓ_1-penalty, which are commonly used for model selection in regression settings (Tibshirani, 1996). Thus, we update the complexity penalty as $P(\tilde{\boldsymbol{\Omega}}) = \|\tilde{\boldsymbol{B}}\|_1$, where $\|A\|_1 = \sum_{j=1}^{p}\sum_{q=1}^{k}|a_{jq}|$, for $a_{jq} \in A$. This penalty not only results in sparse loadings estimates, but also can be used to prevent the non-identifiability of the loadings matrix (Scharf and Nestler, 2019).

Finally, by minimizing the loss function (2.3) as

$$(\hat{\boldsymbol{B}}_{\tilde{k},\lambda}, \hat{\boldsymbol{\Sigma}}_{\tilde{k},\lambda}) \equiv \underset{\tilde{\boldsymbol{B}},\tilde{\boldsymbol{\Sigma}}}{\operatorname{argmin}} \left\{\log|\tilde{\boldsymbol{\Omega}}| + \operatorname{tr}\left(\tilde{\boldsymbol{\Omega}}^{-1}\overline{\boldsymbol{\Omega}}\right) + \lambda\|\tilde{\boldsymbol{B}}\|_1\right\}, \tag{2.4}$$

subject to $\tilde{\boldsymbol{\Omega}} = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{B}}^T + \tilde{\boldsymbol{\Sigma}}$, to different values of $\tilde{k}$ and a given complexity parameter λ we obtain the optimal estimates $\hat{\boldsymbol{B}}_{\tilde{k}}$ and $\hat{\boldsymbol{\Sigma}}_{\tilde{k},\lambda}$, referred as DSSFA estimates, for the factor loadings and uniqueness, respectively.

A direct result from (2.4) implies that the posterior covariance matrix $\overline{\boldsymbol{\Omega}}$ is the optimal solution when $\tilde{k} = k$, and $\lambda = 0$. If there is an interest solely in the number of factors the optimization can be done with $\lambda = 0$ for different values of k. In this paper, the maximum value for $\tilde{k}$ was chosen as the dimension of the full model posterior, k.

There is an efficient off-the-shelf procedure for solving (2.4) when $\lambda > 0$. We used the `fanc` package (Hirose and Yamamoto, 2015; Kei Hirose, 2016) from `R` (R Core Team, 2020) in which we replace the sample covariance by $\overline{\boldsymbol{\Omega}}$ in the package's main function (`fanc`) to perform the optimization. This function use the MC+ penalty (Zhang et al., 2010) as default, which is a non-convex function indexed by `rho` > 0. To obtain the soft threshold, we let `rho` $\to \infty$ in the arguments of the function. Other possible methods to solve (2.4) can be seen in Scharf and Nestler (2019).

The selected optimization method is capable of handling models with p in the thousands (see Hirose and Yamamoto, 2015). In our experience, the method performs well when the number of factors is small in comparison to the number of variables and when there is no regularization on the loadings. In Section 3, there are runtimes from different simulation settings using this optimization method compared to standard Bayesian methods.

2.2 Posterior summary

As seen in Section 2.1, we can assess the information contained in the multivariate posterior of the factor analysis model by using the loss function (2.1). Here, we use the same loss function to generate a posterior summary that measures the trade-off between the entire model and its lower-dimensional representation.

First, we replace the action $\tilde{\boldsymbol{\Omega}}$ in the loss function (2.1) by the DSSFA covariance estimate,

$$\hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda} = \hat{\boldsymbol{B}}_{\tilde{k},\lambda}\hat{\boldsymbol{B}}^{T}_{\tilde{k},\lambda} + \hat{\boldsymbol{\Sigma}}_{\tilde{k},\lambda}, \tag{2.5}$$

where $\hat{\boldsymbol{B}}_{\tilde{k},\lambda}$ and $\hat{\boldsymbol{\Sigma}}_{\tilde{k},\lambda}$ are the optimal estimates obtained from (2.4). In doing so, we generate a sequence of loss functions, $\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})$, see Step 2 of Section 2.3, indexed by the factor dimension $\tilde{k}$ and the complexity parameter λ. This sequence enables us to switch from a complete model with $\tilde{k} = k$ and $\lambda = 0$ to a sparse low-dimensional representation, in which we may more clearly assess the effect of the deterioration in fit. Second, we summarize these changes in a grid, which can be visually inspected in a plot, exposing the trade-off between fit, factor dimension, and sparseness. Importantly, models generated by estimates $\hat{\boldsymbol{B}}_{\tilde{k},\lambda}$, whose columns are zeroed by the optimization procedure, are discarded from the summary.

We also propose a criterion that simultaneously selects the model with the lowest factor dimension and the sparsest loadings matrix while maintaining the full model's fit. We consider the model that generates the greatest expected posterior loss,

$$E_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})] = \log|\hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda}| + \operatorname{tr}\left(\hat{\boldsymbol{\Omega}}^{-1}_{\tilde{k},\lambda}\overline{\boldsymbol{\Omega}}\right), \tag{2.6}$$

that is within a quantile of the full model's loss, $\mathcal{L}_{k,0}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{k,0})$. This in turn provides the maximum acceptable trade-off between the full model produced by the posterior distribution and the DSSFA estimates.

We refer to the factor dimension and complexity parameter selected by this criterion as k^*, and λ^*, respectively. By using this criterion, we are able to select a simple factor analysis model. Moreover, it enables the factor model to be chosen automatically without the need for a visual inspection of the summary plot. In this paper, we consider quantiles between 95% and 99% of the loss function of the full model. This criterion was applied to a toy example in Section 2.4, numerical examples in Section 3, and an empirical application in Section 4.

2.3 Method overview

Before illustrating the proposed approach with an example, we present an overview of the DSSFA method summarized in two steps.

We initiate our procedure by obtaining the expected posterior covariance matrix, $\overline{\boldsymbol{\Omega}} = \overline{\boldsymbol{B}\boldsymbol{B}^T} + \overline{\boldsymbol{\Sigma}}$, from the posterior distributions, $p(\boldsymbol{B}|\boldsymbol{y})$, and $p(\boldsymbol{\Sigma}|\boldsymbol{y})$, with factor dimension set as k. We approximate $\overline{\boldsymbol{B}\boldsymbol{B}^T} \approx \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{B}_{(m)}\boldsymbol{B}_{(m)}^T$ and $\overline{\boldsymbol{\Sigma}} \approx \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\Sigma}_{(m)}$ where $\boldsymbol{B}_{(m)}$ and $\boldsymbol{\Sigma}_{(m)}$ are the posterior samples with $m = 1, 2, \ldots, M$.

Step 1 DSSFA estimates: Apply the posterior mean of the covariance matrix $\overline{\boldsymbol{\Omega}}$, to the optimization procedure in the package `fanc` to solve (2.4). Obtain a sequence of sparse loadings and uniqueness, $(\hat{\boldsymbol{B}}_{\tilde{k},\lambda},\ \hat{\boldsymbol{\Sigma}}_{\tilde{k},\lambda})$, for $\tilde{k} = 1, \ldots, k$, indexed by $\lambda = \lambda_0, \lambda_1, \ldots, \lambda_l$, where $\lambda_0 = 0$ and λ_l is determined by the optimization method (`fanc`) given a choice for the length of the sequence $l \in \{1, 2, \ldots\}$.

Step 2 Summary plot: From the DSSFA estimates obtained in step one, generate the sequence

$$\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda}) = \log|\hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda}| + \operatorname{tr}\left(\hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda}^{-1}\boldsymbol{\Omega}\right), \tag{2.7}$$

for $\tilde{k} = 1, \ldots, k$, and $\lambda = \lambda_0, \lambda_1, \ldots, \lambda_l$. The posterior distribution of $\boldsymbol{\Omega}$ in equation (2.7) is approximated by the posterior samples $\boldsymbol{\Omega}_{(m)}$, with $\boldsymbol{\Omega}_{(m)} = \boldsymbol{B}_{(m)}\boldsymbol{B}_{(m)}^T + \boldsymbol{\Sigma}_{(m)}$, for $m = 1, 2, \ldots, M$. Select a quantile for the loss function of the full model, and plot the expected values $E_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})]$ of (2.7), in relation to $\tilde{k}$ and λ in a graphical summary. The exception can be approximated by replacing the posterior parameter $\boldsymbol{\Omega}$ by $\overline{\boldsymbol{\Omega}}$ as seen in equation (2.6).

Ultimately, it is left to the end user to decide the best lower dimension representation given the quantile of the loss generated by the full model. Otherwise, one may automate the procedure by using the criterion provided in Section 2.1.

2.4 Toy example

In this section, we present a toy example to illustrate the DSSFA approach. We applied our method to simulated data generated from a factor analysis model with known loadings extracted from Harman (1976). Originally, the loadings

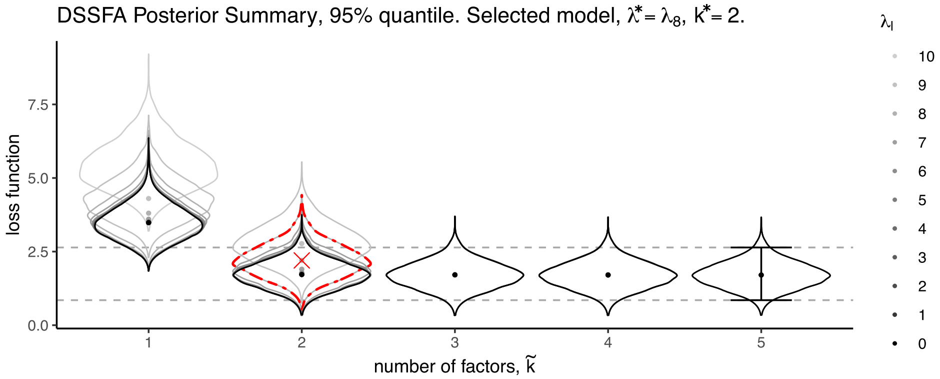

Figure 1: DSSFA summary plot for the toy example, comparing the densities of the loss functions, $\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})$, the violin plots, with respective posterior means $E_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})]$, the dots, indexed by the number of factors $\tilde{k} = 1, 2, \ldots, 5$, and the penalty parameters $\lambda = \lambda_0, \lambda_1, \ldots, \lambda_{10}$, obtained according to Step 1 and Step 2 in the method's overview in Section 2.3. The dashed line is the 95% quantile of the loss function of the full model with no penalty, $\lambda_0 = 0$, identified with an error bar in $\tilde{k} = 5$. From the criterion presented in Section 2.3, we consider the model that generated the loss function (color dot-dashed density), with the greatest expected value, identified as $\times$, that is within the 95% quantile of the loss function of the full model. The resulted loss represents a model with $k^* = 2$ factors, and complexity parameter $\lambda^* = \lambda_8$, which results in a loadings matrix, $\hat{\boldsymbol{B}}_{2,\lambda_8}$, with 19% zeroed entries.

$$\boldsymbol{B}_0 = \begin{pmatrix} 0.879 & 0.919 & 0.890 & 0.858 & 0.238 & 0.183 & 0.135 & 0.250 \\ 0.272 & 0.210 & 0.182 & 0.246 & 0.900 & 0.792 & 0.729 & 0.684 \end{pmatrix}^T,$$

came from the analysis of eight physical variables from 305 individuals, where the number of factors was determined as $k_0 = 2$.

In this example, we generated $n = 100$ samples from model (1.1), with $\boldsymbol{B}_0$ and the uniqueness generated as $\boldsymbol{\Sigma}_0 = \text{diag}(\boldsymbol{I}_p - \boldsymbol{B}_0\boldsymbol{B}_0^T)$, where diag$(A)$ is the matrix formed by the diagonal elements of the matrix A. The normal distribution $N(0, \eta)$ was assigned to the loadings of $\boldsymbol{B}_0$, with η and σ_j, the idiosyncratic errors in $\boldsymbol{\Sigma}_0$, following an Inverse Gamma distribution $(\eta, \sigma_j \sim IG(1, 1))$, for $j = 1, 2, \ldots, p$. No constrains were used for identification. We set the factor dimension as $k = 5$, and ran the Gibbs sampler for 10,000 iterations, discarding the first 5,000 as burn-in. The Gibbs sampler for the factor analysis algorithm is implemented in Rcpp (Eddelbuettel and François, 2011), and is available in the supplementary material of Man and Culpepper (2022). We obtained the posterior mean of the covariance matrix $\overline{\boldsymbol{\Omega}}$ and followed the steps presented in Section 2.3. In the optimization procedure, we let $\tilde{k} = 1, 2, \ldots, 5$, and with penalized solution path of size $\lambda = \lambda_0, \lambda_1, \ldots, \lambda_{10}$, where $\lambda_0 = 0$, for each factor dimension $\tilde{k}$.

Figure 1 displays the DSSFA posterior summary plot, where the 95% quantile of the loss of the full model is shown (dashed line). The increasing values of the loss functions

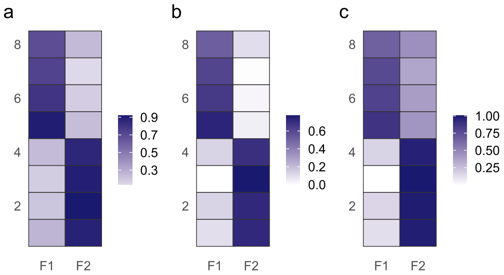

Figure 2: DSSFA estimates of the loading matrix with sparse representation $\hat{\boldsymbol{B}}_{2,\lambda_8}$, Figure (b), and without regularization $\hat{\boldsymbol{B}}_{2,\lambda_0}$, Figure (c), compared to the true loading matrix, $\boldsymbol{B}_0$, in Figure (a).

$\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})$ in relation to λ and $\tilde{k}$ show a deterioration in fit. From the posterior summary, models with $\tilde{k} = 1$ factors were not considered, since the expected posterior losses, $E_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})]$, were not within the 95% quantile of the loss function of the full model. Models with $\tilde{k} \geq 2$ or greater, were considered since the values of the expected posterior losses are inside the specified quantile. Furthermore, we observe a smaller solution path generated by λ for models with $\tilde{k} \geq 3$. This behavior is caused by the fact that the optimization procedure returns loading matrices with zeroed columns even for small values of λ, which are discarded from the summary.

The model highlighted in Figure 1 was selected using the criterion defined in Section 2.2, which returned the complexity coefficient $\lambda^* = \lambda_8$ and the factor dimension $k^* = 2$. This resulted in a loadings matrix with 19% of zeroed entries. Figure 2 displays the recovered loadings matrices $\hat{\boldsymbol{B}}_{2,\lambda_8}$, the penalty-free loadings, $\hat{\boldsymbol{B}}_{2,\lambda_0}$, and $\boldsymbol{B}_0$.

3 Simulation study

In this section, we evaluate the performance of the DSSFA method in the recovery of the true covariance matrix $\boldsymbol{\Omega}_0 = \boldsymbol{B}_0\boldsymbol{B}_0^T + \boldsymbol{\Sigma}_0$. We compare our method with the marginal likelihood estimate for selecting the number of factors (Lopes and West, 2004; Newton and Raftery, 1994), and BMA (Hoeting et al., 1999) for covariance matrix estimation.

We provide evidence that DSSFA enhances factor dimension selection without significantly impacting the estimation of $\boldsymbol{\Omega}_0$, under different simulation settings. When compared with the standard Bayesian procedure, our method also shows substantial running time gains in scenarios with significantly large dimensions.

3.1 Simulation settings

In this study, we used three different simulation settings. In setting 1, we followed the simulation design of Man and Culpepper (2022), in which the synthetic data was drawn from model (1.1), with sample sizes of $n = 100, 500$, and 1000, with $p = 15$ variables and

$k_0 = 3$ factors. The entries of the loadings matrix $\boldsymbol{B}_0$ were independently sampled from a standard normal distribution $N(0, 1)$ for each replicate. The idiosyncratic matrix was set as $\boldsymbol{\Sigma}_0 = \sigma^2 \boldsymbol{I}_p$, with $\sigma = 0.2$, and 0.5. In setting 2, we investigated the robustness of our approach by employing the standard multivariate t-distribution, $t_{k_0}(0, \boldsymbol{I}_{k_0}, \nu)$ for $\boldsymbol{f}_i$, and the scaled t-distribution $t_p(\mathbf{0}, \boldsymbol{\Sigma}_0, \nu)$ for $\boldsymbol{\epsilon}_j$, with degrees of freedom of $\nu = 3$, and 10. Under this setup, we evaluated our method on data of size $n = 100$. The variance of the idiosyncratic errors were generated as in setting 1. In setting 3, we explored our method under the normal standard factor model of setting 1, with different dimensions and number of factors. We selected two challenging scenarios: (i) $p = 50$ variables, $k_0 = 5$ factors, and sample size $n = 100$, and (ii) $p = 100$ variables, $k_0 = 10$ factors, and $n = 500$ samples. In this setting, the idiosyncratic errors, σ, were generated by a uniform distribution with ranges between 0.5 and 0.8 in each interaction. In all settings the resulting loadings were rotated to be PLT to assure identifiability.

We included three different priors in the study. The first is from Geweke and Zhou (1996), which uses the standard PLT constraint on the loadings. We refer to this prior as GZ. The second, is a novel prior presented by Man and Culpepper (2022) which incorporates a PLT type constraint with a mode-jumping step to avoid multimodal posteriors. This prior is referred to as MC. The third is the plain normal prior on the loadings without constraints, which we refer as unconstrained (UN).

We ran Monte Carlo Markov Chain (MCMC) algorithms for 30,000 iterations discarding the first 15,000 as burn-in in all settings. In settings 1, and 2, for each algorithm, we generated posteriors with factor dimensions $k = 1, 2, \ldots, 5$. For simulation setting 3, we generated posteriors with factor dimensions $k = 1, 2, \ldots, 10$ in scenario (i), and $k = 1, 2, \ldots, 15$ in scenario (ii). We generated the posterior samples from priors using algorithms from the supplementary material of Man and Culpepper (2022). We ran 300 replicates for each simulation setting on an Intel Core i5 CPU laptop computer with 7.7 GB of RAM.

Prior specifications

We followed the prior and hyperparameter specifications used in the study of Man and Culpepper (2022). In the algorithm from Geweke and Zhou (1996) the loadings were assigned independent normal priors $b_{jk0}|\eta \sim N(0, \eta)$, with $b_{jk} \in \boldsymbol{B}_0$, for $j = 1, \ldots, p$, and k the fixed number of factors, where $b_{jk} = 0$ for $k > j$, and $b_{kk} > 0$. Man and Culpepper (2022) relaxed the standard constraint of Geweke and Zhou (1996), and applied the PLT constraint to any arbitrary subset of the p rows of the loadings matrix. Hence, for a permutation set $\boldsymbol{r} = (r_1, r_2, \ldots, r_k) \subset \{1, 2, \ldots, p\}$, with $r_k \neq r_{k'}$, and $k \neq k'$, the loadings were constrained as $b_{r_k k'} = 0$ for $k' > k$, and $b_{r_k k} > 0$. To avoid near singular cases in the sub matrix generated by $\boldsymbol{r}$, Man and Culpepper (2022) used the prior distribution, $p(b_{r_k k}|\eta, \gamma) \propto b_{r_k k}^{\gamma} \exp(-b_{r_k k}^2 / 2\eta)$, for $b_{r_k k} > 0$, where $\gamma = 0.5$. A sample from the non-singular restricted PLT submatrix is obtained with a Metropolis-Hastings step, which is incorporated into the loadings matrix via matrix decomposition (see Man and Culpepper, 2022). Lastly, an uniform prior was assigned to the permutation set $\boldsymbol{r}$. The normal distribution $N(0, \eta)$ was assigned to the remaining unrestricted loadings. In the Unconstrained model, the entirety of loadings was assigned the normal distribution $N(0, \eta)$ prior. In all models, $\eta \sim IG(1, 1)$ was chosen for the variance of the loadings, and $\sigma_j \sim IG(1, 1)$, for $j = 1, 2, \ldots, p$, for the idiosyncratic variances.

3.2 Evaluation and results

We evaluated the DSSFA method uniquely on the posterior samples generated with a factor dimension of size $k = 5$ in simulation settings 1 and 2. In simulation setting 3, we used our method on the posterior samples with factor dimensions of size $k = 10$ and $k = 15$, yielded by scenarios (i) and (ii), respectively. In the optimization step, we generated estimates with $\tilde{k} = 1, 2, \ldots, 5$ factors for simulation settings 1 and 2, and $\tilde{k} = 1, 2, \ldots, 10$, for scenario (i), and $\tilde{k} = 1, 2, \ldots, 15$ for scenario (ii) of simulation setting 3. We set the complexity parameter as $\lambda = 0$ in all simulation settings, since we are not interested in the sparsity of the loadings in this study. We used the criterion described in Subsection 2.3, and let the method auto-select point estimates under 95% and 99% quantiles of the loss of the full model. As a result, we obtained the factor dimension k^* and the DSSFA estimates $\hat{\boldsymbol{B}}_{k^*}$ and $\hat{\boldsymbol{\Sigma}}_{k^*}$, from which we recovered the covariance matrix estimate $\hat{\boldsymbol{\Omega}}_{k^*}$, as seen in equation (2.5).

In contrast to our procedure, we used bridge sampling from the `R` package presented in Gronau et al. (2020) to calculate the marginal likelihood estimate for each posterior dimension in all simulation settings to identify the number of factors. We recall that Lopes and West (2004) and Man and Culpepper (2022) favor bridge sampling to other techniques like the harmonic mean and Newton-Raftery's estimators (Newton and Raftery, 1994). From the same marginal likelihood estimates, we obtained the weights and recovered the BMA estimates of the covariance matrix.

In order to compare the two methodologies, we evaluated the percentage of correctly identified factor dimensions in each simulation setting. To evaluate the recovery of the covariance estimate, we used the root mean squared error (RMSE), which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{p^2}\sum_{j=1}^{p}\sum_{q=1}^{p}(\hat{\boldsymbol{\Omega}}_{jq} - \boldsymbol{\Omega}_{0jq})^2}, \quad \boldsymbol{\Omega}_{0jq} \in \boldsymbol{\Omega}_0, \quad \hat{\boldsymbol{\Omega}}_{jq} \in \hat{\boldsymbol{\Omega}}, \tag{3.1}$$

where $\hat{\boldsymbol{\Omega}}$ is the recovered point estimates from the two procedures, and $\boldsymbol{\Omega}_0$ is the true covariance matrix.

Additionally, in simulation setting 3, we timed how long the studied procedures ran. For the DSSFA method, we took into account the MCMC method's runtimes for the largest dimension and added the optimization runtime. For the marginal likelihood approach, all the MCMC posterior dimensions along with the bridge sampling runtime, were taken into account.

Simulation results

Table 1 on page 192 shows the percentage of the models that were accurately identified using both methods. In simulation setting 1, under the GZ prior, the bridge sampling-

Prior		GZ			UN			MC		
Method		DSSFA		ML	DSSFA		ML	DSSFA		ML
Quantile		95%	99%		95%	99%		95%	99%	
setting 1	(n, σ, k_0)									
	$(100, 0.2, 3)$	100	100	71	100	100	100	100	100	100
	$(100, 0.5, 3)$	100	100	80	100	100	100	100	100	100
	$(500, 0.2, 3)$	100	100	78	100	100	100	100	100	100
	$(500, 0.5, 3)$	100	100	86	100	100	100	100	100	100
	$(1000, 0.2, 3)$	100	100	84	100	100	100	100	100	100
	$(1000, 0.5, 3)$	100	100	84	100	100	100	100	100	100
setting 2	(ν, σ, k_0)									
	$(3, 0.2, 3)$	70	80	40	74	82	52	64	76	44
	$(3, 0.5, 3)$	61	74	23	68	80	32	61	72	30
	$(10, 0.2, 3)$	100	100	80	100	100	100	100	100	100
	$(10, 0.5, 3)$	100	100	81	100	100	100	100	100	100
setting 3	(n, p, k_0)									
	$(100, 50, 5)$	88	100	89	95	100	100	77	100	100
	$(500, 100, 10)$	93	100	67	100	100	0	100	100	100

Table 1: Percentage of correctly identified models in 300 replications for the three simulation settings. In simulation setting 1, we have the recovered proportions of a normal factor model with $n = 100, 500$, and 1000, with $\sigma = 0.2$, and 0.5, $p = 15$, and factor dimension of $k_0 = 3$. In setting 2, we have the recovered proportions of a t-distributed factor model with $\nu = 3$, and 10, standard deviations $\sigma = 0.2$, and 0.5, with sample size $n = 100$, $p = 15$, and factor dimension $k_0 = 3$. In setting 3, we have the recovered proportions of a the normal factor model with $n = 100$, $p = 50$, and $k_0 = 5$ in scenario (i), and $n = 500$, $p = 100$, and $k_0 = 10$ in scenario (ii). The number of factors were identified using the DSSFA method with 95% and 99% quantiles of the loss function of the full model, and for the marginal likelihood estimate (ML) we used bridge sampling. We considered the Geweke & Zhou (GZ), Unconstrained (UN) and Man & Culpepper (MC) priors in the simulation.

based marginal likelihood estimate selected the incorrect number of factors in every scenario. According to Man and Culpepper (2022), the posterior multimodality produced by the PLT constraint may be a major contributor to much of the uncertainty in the marginal likelihood estimation. the DSSFA method overcomes this issue by selecting the correct number of factors in 300 out of 300 replicates using both quantiles in all situations under this prior. Furthermore, our approach correctly identified the number of factors for the MC and UN priors in each and every replicate for both quantiles in every scenario of this setup. The same outcome was obtained with the marginal likelihood estimate.

For data generated using t-distributed factors and errors with $\nu = 3$ degrees of freedom in simulation setting 2, the DSSFA approach outperformed the marginal likelihood estimate for all scenarios and priors. The 99% quantile choice makes this discrepancy more apparent. This outcome might be caused by the fact that the resulting loss function, (2.7), reflects the noise in the data and is therefore more scattered, making a larger interval more efficient. We also noticed that the DSSFA and marginal likelihood estimates performed similarly when applied to the data generated in simulation setting 1

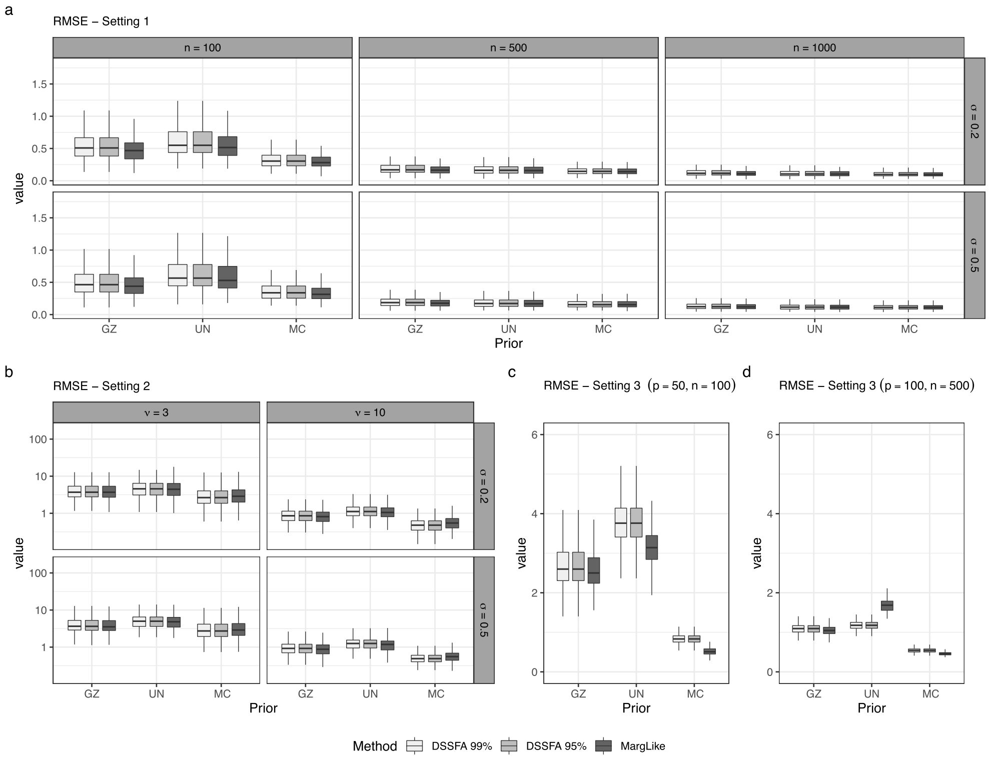

Figure 3: RMSE from 300 replicates, between the true covariance matrix $\mathbf{\Omega}_0$ and the estimates generated by the DSSFA method, with 95% and 99% quantiles, and Bayes model averaging (MargLike), with priors: Geweke & Zhou (GZ), Unconstrained (UN), and Man & Culpepper (MC) in different simulation settings. Figure 1 (a), displays the results from simulation setting 1, given a normal factor model with $n = 100, 500$, and 1000, $\sigma = 0.2$, and 0.5 and true factor dimension $k_0 = 3$. Figure 2 (b), displays the log scaled results for simulation setting 2, given a t-student factor model with standard deviations $\sigma = 0.2$, and 0.5, with sample size $n = 100$, and true factor dimension $k_0 = 3$. Figure 2 (c), and Figure 2 (d) displays simulation setting 3 results, with estimates from normal factor model with $n = 50$, $p = 100$, and $k_0 = 5$ in the scenario (i), and $n = 500$, $p = 100$, and $k_0 = 10$ in scenario (ii), respectively.

when we increased the degrees of freedom from 3 to 10.

In simulation setting 3, scenario (i), our method performed better using the 99% quantile than the 95%, in which the true factor dimension was selected 300 out of 300 times under all priors. The marginal likelihood estimates under the MC prior yielded a similar outcome. In this scenario, once again, the GZ prior with marginal likelihood estimate produced the incorrect number of factors. In scenario (ii), the UN prior performed noticeably worse under marginal likelihood estimation with no correct estimation. This

Prior		Geweke & Zhou		Unconstrained		Man & Culpepper	
Method		DSSFA	ML	DSSFA	ML	DSSFA	ML
setting	(n, p, k_0)						
(i)	$(100, 50, 5)$	30.5	385.11	32.5	462.03	25.5	423.63
		(2.15)	(6.74)	(5.78)	(35.13)	(1.73)	(10.87)
(ii)	$(500, 100, 10)$	302.85	3607.78	411.10	5022.02	255.77	3028.70
		(32.34)	(457.62)	(32.5)	(462.03)	(25.5)	(423.63)

Table 2: The average and (standard deviation) runtimes in seconds over 300 replicates of the DSSFA method compared to the marginal likelihood estimate (ML) with bridge sampling from settings (i), with $n = 100$, $p = 50$, and $k_0 = 5$, and (ii) with $n = 500$, $p = 100$, and $k_0 = 10$ from simulation setting 3. The DSSFA method has the total runtime of the optimization procedure added to the maximum dimension of the MCMC method.

might be the result of the loadings having too many parameters for the bridge sampling approach to converge properly (Gronau et al., 2020). With flawless model identification in both quantiles with this prior, the DSSFA method solves this problem. This result was repeated with MC and GZ priors when using the 99% quantile. Once more, GZ prior with marginal likelihood estimate provided inaccurate estimations in this setting. Additionally, we found no discernible changes in the simulation results between this setting and the settings with homogeneous error.

Figure 3 displays the RMSE of the recovery of the true covariance matrix in the same replicates as those used for factor dimension estimation. According to the findings, there were no discernible fit differences between the BMA and the DSSFA estimates. Additionally, the DSSFA estimates preserved properties from the posterior distributions, as can be seen in all cases. The DSSFA estimates recovered under the MC prior outperform both GZ and the UN in terms of fit in all scenarios under this setting.

Figure 3 (a) shows that under simulation setting 1, as sample size grows, our technique effectively estimates $\mathbf{\Omega}_0$. The log scale plot of Figure 3 (b) shows that the RMSE significantly increased across the different priors compared to the other settings, reflecting the t-distributed data used in this setting. Nevertheless, the DSSFA approach and the marginal likelihood estimation both fit the data similarly. The RMSE under the UN prior in Figure 3 (d) under simulation setting 3, scenario (ii), highlighted the problem with the bridge sampling convergence as seen previously, leading to a poor fit.

The overall runtime for the methods from simulation setting 3, scenarios (i), and (ii) are shown in Table 2 on page 194. Under all priors, we observe considerable improvements over marginal likelihood utilizing bridge sampling and with the MCMC running in all dimensions. The DSSFA method was ten to twelve times faster. This outcome is partly attributable to the DSSFA method's reliance on the information available in the posterior generated uniquely by the full model, and the selected optimization technique. Moreover, once the series of loss functions is obtained, no additional runtime is required to generate the different models given a the chosen quantile.

4 Real data analysis

This section aims to show the flexibility and usefulness of our method through an application to personality traits data. We applied the DSSFA method in posterior samples generated from over-fitted priors for factor analysis and obtained an estimate for the number of factors. These priors start with a conservative factor dimension and remove components by shrinking their loadings to zero. The selection of the number of factors is subsequently made using adaptive Gibbs sampling methods. Such approaches included in this category are the multiplicative gamma process (MGP) from Bhattacharya and Dunson (2011), and a novel procedure that uses cumulative shrinkage priors (CUSP), from Legramanti et al. (2020). As seen in Section 2.1, we can use our approach in such models as the DSSFA method has no restrictions on prior choice since it depends uniquely on the posterior distribution of the covariance matrix.

4.1 Personality traits data

For the data analysis, we followed Legramanti et al. (2020) and explored a subset of the big five personality traits data, which is available at `bfi` in the `R` package `psych` (Revelle et al., 2018). We examined the association structure among $p = 25$ personality variables collected from $n = 126$ individuals over age 50. We also centered the data and changed the sign of the variables 1, 9, 10, 11, 12, 22 and 25.

Prior specifications

In the MGP prior, the loadings are distributed as $N(0, \phi_{jk}^{-1}\theta_k^{-1})$, for $j = 1, \dots, p$ and $k \in \{1, 2, \dots\}$, with $\phi_{jk} \sim Gamma(3/2, 3/2)$. For the global precisions θ_k^{-1} we have the multiplicative gamma process prior $\theta_k = \vartheta_1 \cdots \vartheta_k$, with $\vartheta_1 \sim Gamma(2.1, 1)$ and $\vartheta_l \sim Gamma(3.1, 1)$, for $l \geq 2$ and $k \in \{1, 2, \dots\}$. The CUSP prior induces increasing shrinkage via a sequence of spike-and-slab distributions that assign growing mass to the spike as the model complexity grows. Under this prior, the factor loadings $b_{jk} \in \boldsymbol{B}$ are distributed as $N(0, \theta_k)$ for $k \in \{1, 2, \dots\}$, with θ_k assuming the spike and slab mixture $\theta_k|\pi_k \sim (1 - \pi_k)IG(2, 2) + \pi_k \delta_{\theta_\infty}$, where $\pi_k = \sum_{l=1}^{k} \omega_l$, with $\omega_l = \nu_l \prod_{m=1}^{l-1}(1 - \nu_m)$, and $\nu_l \sim Beta(1, 5)$. The spike is defined on δ_{θ_∞}, which is the point mass on θ_∞, with $\theta_\infty = 0.05$. For the two methods, we have $\sigma_j^2 \sim IG(1, 0.3)$, for $j = 1, \dots, p$, for the idiosyncratic variances. The adaptation $p(t) = \exp(\alpha_0 + \alpha_1 t)$ was allowed only after $t = 500$ iterations and (α_0, α_1) were set to $(-1, -5 \times 10^{-4})$, while the adaptation threshold ϵ in the MGP was set as 10^{-4}. The number of factors were initialized as $k = p$ for MGP, and as $k = p + 1$ for CUSP. The MGP method was sampled using the `R` package `infinitefactor`, and CUSP was generated from the algorithm at `https://github.com/siriolegramanti/CUSP`.

Application results

We ran the MCMC algorithms for 10,000 iterations discarding the first 5,000 as burn-in and with thinning in every five samples for both methods. At a first run, MGP obtained

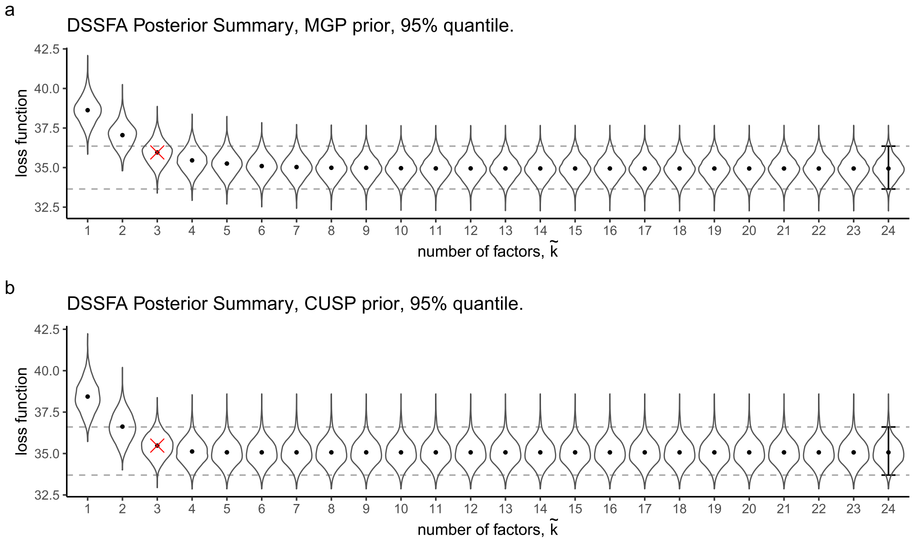

Figure 4: DSSFA posterior summary plots generated with (a) MGP, and (b) CUSP priors, obtained according to the steps presented in Section 2.3. The plots display the densities of the loss functions $\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})$, the violin plots, with respective posterior means $E_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathcal{L}_{\tilde{k},\lambda}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Omega}}_{\tilde{k},\lambda})]$, the dots, indexed by the number of factors $\tilde{k} = 1, 2, \ldots, 24$. No penalty was used. The dashed line is the 95% quantile of the loss function of the full model generated with $\tilde{k} = 24$ (error bar). Following the criterion in Section 2.3, we consider the model that generates a loss function whose expected value, identified as ×, is within the selected quantile of the loss function of the full model. The resulted loss function yields a model with $k^* = 3$ factors, for the MGP, and for CUSP priors.

a posterior mean (95% credible interval) of 20.7 ([18,24]) for the number of factors, while CUSP obtained 2.64 ([2,3]). From the same posterior samples generated by MGP and CUSP, we followed the steps presented in Section 2.3, and applied the DSSFA method with $\tilde{k} = 1, 2, \ldots, (p - 1)$ factors for the two posterior distributions. This upper limit is imposed by the optimization procedure (`fanc`). We let our method auto-select the factor dimension under the 95% quantile of the loss function of the full model and we set the complexity parameter as $\lambda = 0$ in both settings, since we are not interested in the sparsity of the loadings in this study.

Figure 4 displays the posterior summary plots for the two methods. Under a 95% quantile of the loss function of the full model, our method selected a factor model of size $k^* = 3$ for both MGP and CUSP. These results are in agreement with the analysis of Legramanti et al. (2020), in which three main factors were identified. Further analysis, on the posterior summary plots indicate that the MGP posterior contains information of a three-factor model, although the posterior adaptation procedure privileges a model with redundant factors.

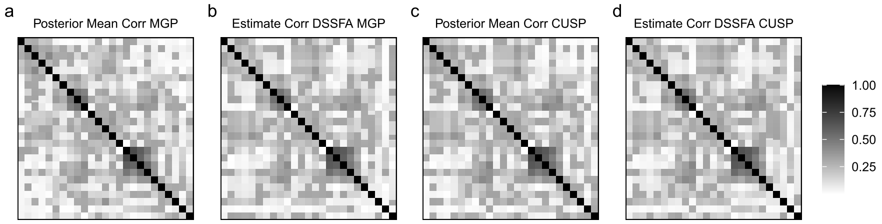

Figure 5: Absolute values of the posterior mean of the correlation matrix under the (a) MGP, and (c) CUSP priors, compared with the absolute values of the correlation matrix estimates produced by the DSSFA method with 95% quantile of the loss function of the full model, under (b) MGP, and (d) CUSP priors.

Figure 5 displays the absolute values of the posterior mean of the correlation matrix, from priors MGP, (a), and CUSP, (c), in comparison to the absolute values of the correlation matrix generated by the DSSFA estimates with the 95% quantile of the loss function of the full model under (b) MGP and (d) CUSP. We observe that the values of the absolute correlations are similar across the different estimates, which indicates that MGP overestimates the number of factors, further confirming the analysis of Legramanti et al. (2020), and the results from the DSSFA posterior summary plot in Figure 4.

4.2 Interpretable loadings from overfitted factor models

In the factor analysis literature, over-fitted factor models are usually used for covariance matrix estimation, and thus there is no need to focus on the identifiability and interpretation of the resulting factor loadings structure. We went a step further in our analysis, and used DSSFA with penalty on the loadings, and same posterior samples generated by the CUSP prior in the previous study, to recover an sparse loadings matrix and to prevent the non-identifiability of the loadings (see Scharf and Nestler, 2019).

We ran our method with $\tilde{k} = 1, 2, \ldots, (p-1)$ factors, with a penalized solution path of size $\lambda = \lambda_0, \lambda_1, \ldots, \lambda_{10}$, for each factor dimension $\tilde{k}$. The DSSFA posterior summary plot can be seen in Figure 6. Under a 95% quantile of the loss of the full model, our method selected a factor analysis model with dimension $k^* = 3$, and complexity parameter $\lambda^* = \lambda_8$. Models with $\tilde{k} \leq 2$ were not considered. We included in our analysis models with $\tilde{k} = 4$ and $\tilde{k} = 5$ factors, since they are within the 95% quantile of the loss function of the full model.

Figure 7 displays the selected loadings, in which we can observe a similar factor structure across the different dimensions, although some of the entries of the penalized models are shrunken in comparison with the entries from the models with no penalty ($\lambda_0 = 0$). As in Legramanti et al. (2020), we noticed significant correlation between agreeableness (A) and extraversion (E) in factor F_2, some correlation between conscientiousness (C) and neuroticism (N) in factor F_1. Lastly, Openness (O) presented less evident weights and almost no association between traits. Furthermore, the small values

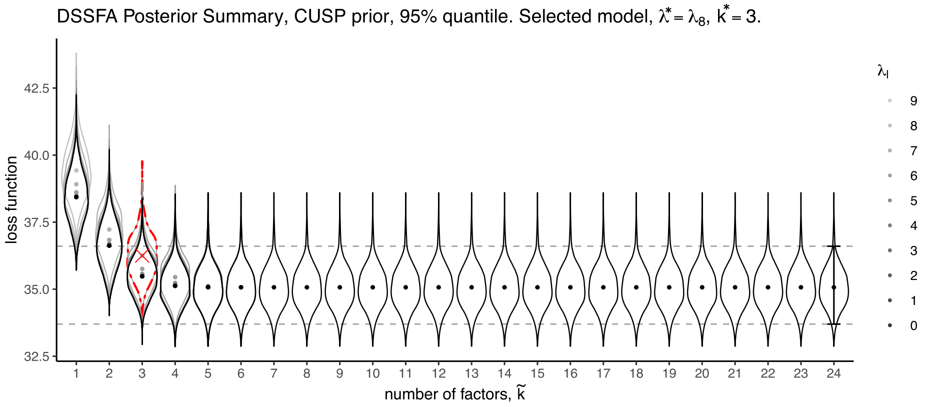

Figure 6: DSSFA summary plot for personality traits data under the CUSP prior comparing the densities of the loss functions $\mathcal{L}_{\tilde{k},\lambda}(\mathbf{\Omega}, \hat{\mathbf{\Omega}}_{\tilde{k},\lambda})$, the violin plots, with respective posterior means $E_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathcal{L}_{\tilde{k},\lambda}(\mathbf{\Omega}, \hat{\mathbf{\Omega}}_{\tilde{k},\lambda})]$, the dots, indexed by the number of factors $\tilde{k} = 1, 2, \ldots, 24$, and the complexity parameters $\lambda = \lambda_0, \lambda_1, \ldots, \lambda_{10}$, obtained according to Step 1 and Step 2, in the method's overview in Section 2.3. The dashed line is the 95% quantile of the loss function of the full model (error bar), with $\tilde{k} = 24$, and with regularization $\lambda_0 = 0$. Following the DSSFA method criterion presented in Subsection 2.3, we consider the model that generates the loss function (color dot-dashed density), with the greatest expected value, identified as $\times$, that is within the 95% quantile of the loss function of the full model. The resulted loss yields a model with $k^* = 3$ factors, and parsimony parameter $\lambda^* = \lambda_8$, which results in a loadings with 39% zeroed entries.

of the loadings in Openness may be associated with the chosen subset of the data (age > 50), suggesting that this trait is less present in the considered age group.

5 Conclusions

In this paper, we presented the DSSFA method, which adds to the literature on factor analysis by defining posterior summarization as a decision problem. The proposed method has two steps: optimization of a predefined loss function and a posterior summary plot. Unlike traditional information-based approaches and hard thresholding rules, our method adds the capability of selecting the factor dimension and sparse loadings simultaneously. Furthermore, given that posterior samples of the Bayesian factor analysis model are available, our method may be used in combination with any prior distribution that is best suited for the situation.

The posterior summary plots revealed the relationship between posterior uncertainty, sparsity and predictive degradation. From these relations, we proposed a criterion that automates the problem of determining the number of factors and the complexity parameter. We performed an extensive simulation study based on this criterion that provided

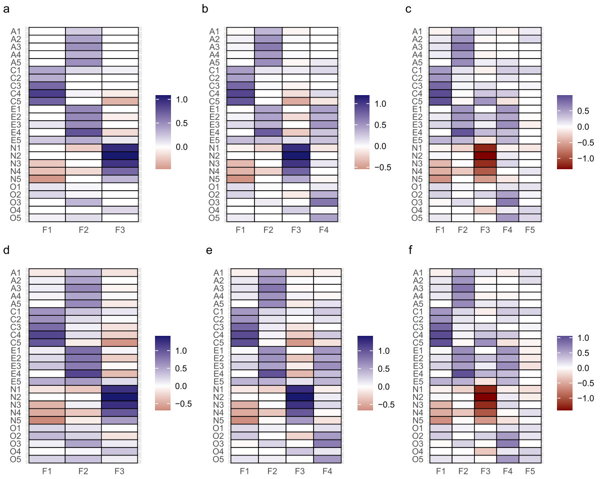

Figure 7: DSSFA estimates under CUSP prior for the personality traits data. Figures (a) and (d) show the resulting sparse loadings $\hat{\boldsymbol{B}}_{3,\lambda_8}$, and loadings without penalty $\hat{\boldsymbol{B}}_{3,\lambda_0}$ selected with the DSSFA procedure. We also include in our analysis models with $\tilde{k} = 4$ factors with penalty $\lambda = \lambda_7$, $\hat{\boldsymbol{B}}_{4,\lambda_7}$ in (b), and without penalty, $\hat{\boldsymbol{B}}_{4,\lambda_0}$ in (e), and models with $\tilde{k} = 5$ factors with penalty $\lambda = \lambda_5$, $\hat{\boldsymbol{B}}_{5,\lambda_5}$ in (c), and without penalty $\hat{\boldsymbol{B}}_{5,\lambda_0}$ in (f). The presented models are within the 95% quantile interval of the loss function of the full model, see Figure 6. The percentage of zeroed loadings in the regularized models in (a), (b), and (c) are 39%, 27%, and 20%, respectively.

evidence of model selection improvement and significant gains in runtime over other widely used procedure.

The usefulness of our method was further assessed by uncovering redundant factors in over-fitted factor models in the application. It was shown that our approach offers an effective alternative for obtaining interpretable loading matrices from models when the posterior samples have different factor dimensions. In this setting, the DSSFA method contributes to recent literature that suggests avoiding the factor model identifiability issue and obtaining point estimates of the posterior distribution either through post-processing the MCMC chains (Papastamoulis and Ntzoufras, 2022; Poworoznek et al., 2021) or by choosing the maximum a posteriori (Schiavon et al., 2022). However, one downside of our approach is the overshrinkage of the loadings in the recovered sparse models.

Future research should look into loss functions other than the negative log-likelihood. Viable options include squared loss, and the Frobenius distance. Other criteria for model selection can also be easily explored. In applications with a large number of variables, such as genetic data, the optimization procedure could run increasingly and stop when the expected loss is within the quantile of the full model's loss. A natural extension of this criterion would be an application to high-dimensional data. Furthermore, other penalties could be used to circumvent the problem of overshrinkage of the loadings observed in the toy example and application. A possible alternative is the Bayesian adaptive penalty presented recently by Kowal et al. (2021). Finally, we envisage the extension of the DSSFA approach to factor regression (West, 2003), in dynamic factor models (Nakajima and West, 2013; Kastner, 2019), and other latent variable models (Bartholomew et al., 2011).

References

Akaike, H. (1987). "Factor analysis and AIC." *Psychometrika*, 52(3): 317–332. MR0914459. doi: https://doi.org/10.1007/BF02294359. 182

Bartholomew, D. J., Knott, M., and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*, volume 904. John Wiley & Sons. MR2849614. doi: https://doi.org/10.1002/9781119970583. 200

Bashir, A., Carvalho, C. M., Hahn, P. R., and Jones, M. B. (2019). "Post-processing posteriors over precision matrices to produce sparse graph estimates." *Bayesian Analysis*, 14(4): 1075–1090. MR4044846. doi: https://doi.org/10.1214/18-BA1139. 182, 184

Berger, J. (2013). *Statistical Decision Theory: Foundations, Concepts, and Methods*. Springer Science & Business Media. MR0580664. 184

Bernardo, J. M. and Smith, A. F. (2009). *Bayesian Theory*, volume 405. John Wiley & Sons. MR1274699. doi: https://doi.org/10.1002/9780470316870. 181, 184

Bhattacharya, A. and Dunson, D. B. (2011). "Sparse Bayesian infinite factor models." *Biometrika*, 98(2): 291–306. MR2806429. doi: https://doi.org/10.1093/biomet/asr013. 182, 183, 195

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008). "High-dimensional sparse factor modeling: applications in gene expression genomics." *Journal of the American Statistical Association*, 103(484): 1438–1456. MR2655722. doi: https://doi.org/10.1198/016214508000000869. 182, 185

Carvalho, C. M., George, E. I., Hahn, P. R., and McCulloch, R. E. (2021). "Variable Selection and Interaction Detection with Bayesian Additive Regression Trees." In *Handbook of Bayesian Variable Selections*, 117–154. Chapman and Hall/CRC. 182

Dey, D. K. and Srinivasan, C. (1985). "Estimation of a covariance matrix under Stein's loss." *The Annals of Statistics*, 13(4): 1581–1591. MR0811511. doi: https://doi.org/10.1214/aos/1176349756. 184

Eddelbuettel, D. and François, R. (2011). "Rcpp: Seamless R and C++ Integration." *Journal of Statistical Software*, 40(8): 1–18. 188

Fruehwirth-Schnatter, S. and Lopes, H. F. (2018). "Sparse Bayesian factor analysis when the number of factors is unknown." *arXiv preprint* arXiv:1804.04231. 181, 182

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC. MR2260716. 183

Geweke, J. and Zhou, G. (1996). "Measuring the pricing error of the arbitrage pricing theory." *The review of financial studies*, 9(2): 557–587. 183, 190

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). "bridgesampling: An R Package for Estimating Normalizing Constants." *Journal of Statistical Software*, 92(10): 1–29. 191, 194

Hahn, R. P. and Carvalho, C. M. (2015). "Decoupling Shrinkage and Selection in Bayesian Linear Models: A Posterior Summary Perspective." *Journal of the American Statistical Association*, 110(509): 435–448. MR3338514. doi: https://doi.org/10.1080/01621459.2014.993077. 182, 184

Harman, H. H. (1976). *Modern Factor Analysis*. University of Chicago press. MR0400546. 187

Hirose, K. and Yamamoto, M. (2015). "Sparse estimation via nonconcave penalized likelihood in factor analysis model." *Statistics and Computing*, 25(5): 863–875. MR3375622. doi: https://doi.org/10.1007/s11222-014-9458-0. 186

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). "Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and EI George, and a rejoinder by the authors." *Statistical science*, 14(4): 382–417. MR1765176. doi: https://doi.org/10.1214/ss/1009212519. 182, 189

Huber, F., Koop, G., and Onorante, L. (2021). "Inducing Sparsity and Shrinkage in Time-Varying Parameter Models." *Journal of Business & Economic Statistics*, 39(3): 669–683. MR4272927. doi: https://doi.org/10.1080/07350015.2020.1713796. 182

Kastner, G. (2019). "Sparse Bayesian time-varying covariance estimation in many dimensions." *Journal of Econometrics*, 210(1): 98–115. MR3944765. doi: https://doi.org/10.1016/j.jeconom.2018.11.007. 182, 200

Kei Hirose, H. N., Michio Yamamoto (2016). *fanc: Penalized Likelihood Factor Analysis via Nonconvex Penalty.*. R package version 2.2. 186

Kowal, D. R. (2021). "Fast, Optimal, and Targeted Predictions Using Parameterized Decision Analysis." *Journal of the American Statistical Association*, 0(0): 1–12. 184

Kowal, D. R. and Bourgeois, D. C. (2020). "Bayesian function-on-scalars regression for high-dimensional data." *Journal of Computational and Graphical Statistics*, 29(3): 629–638. MR4153187. doi: https://doi.org/10.1080/10618600.2019.1710837. 182

Kowal, D. R., Bravo, M., Leong, H., Bui, A., Griffin, R. J., Ensor, K. B., and Miranda, M. L. (2021). "Bayesian variable selection for understanding mixtures in environmental exposures." *Statistics in medicine*, 40(22): 4850–4871. MR4315456. doi: https://doi.org/10.1002/sim.9099. 200

Krantsevich, C., Hahn, P. R., Zheng, Y., and Katz, C. (2021). "Bayesian decision theory for tree-based adaptive screening tests with an application to youth delinquency." *arXiv preprint* arXiv:2106.10364. 182

Kullback, S. (1997). *Information Theory and Statistics*. Courier Corporation. 184

Legramanti, S., Durante, D., and Dunson, D. B. (2020). "Bayesian cumulative shrinkage for infinite factorizations." *Biometrika*, 107(3): 745–752. MR4138988. doi: https://doi.org/10.1093/biomet/asaa008. 182, 183, 195, 196, 197

Lopes, H. F. and West, M. (2004). "Bayesian model assessment in factor analysis." *Statistica Sinica*, 14(1): 41–67. MR2036762. 182, 183, 189, 191

Man, A. X. and Culpepper, S. A. (2022). "A mode-jumping algorithm for Bayesian factor analysis." *Journal of the American Statistical Association*, 117(537): 277–290. MR4399085. doi: https://doi.org/10.1080/01621459.2020.1773833. 182, 183, 188, 189, 190, 191, 192

Nakajima, J. and West, M. (2013). "Bayesian analysis of latent threshold dynamic models." *Journal of Business & Economic Statistics*, 31(2): 151–164. MR3055329. doi: https://doi.org/10.1080/07350015.2012.747847. 182, 200

Newton, M. A. and Raftery, A. E. (1994). "Approximate Bayesian inference with the weighted likelihood bootstrap." *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1): 3–26. MR1257793. 182, 189, 191

Papastamoulis, P. and Ntzoufras, I. (2022). "On the identifiability of Bayesian factor analytic models." *Statistics and Computing*, 32(2): 1–29. MR4394853. doi: https://doi.org/10.1007/s11222-022-10084-4. 183, 199

Poworoznek, E., Ferrari, F., and Dunson, D. (2021). "Efficiently resolving rotational ambiguity in Bayesian matrix sampling with matching." *arXiv preprint* arXiv:2107.13783. 199

Puelz, D., Hahn, P. R., Carvalho, C. M., et al. (2017). "Variable selection in seemingly unrelated regressions with random predictors." *Bayesian Analysis*, 12(4): 969–989. MR3724975. doi: https://doi.org/10.1214/17-BA1053. 182

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 186

Revelle, W. et al. (2018). "psych: Procedures for psychological, psychometric, and personality research." *R package version*, 1(10). 195

Ročková, V. and George, E. I. (2016). "Fast Bayesian factor analysis via automatic rotations to sparsity." *Journal of the American Statistical Association*, 111(516): 1608–1622. MR3601721. doi: https://doi.org/10.1080/01621459.2015.1100620. 181

Scharf, F. and Nestler, S. (2019). "Should regularization replace simple structure rotation in exploratory factor analysis?" *Structural Equation Modeling: A Multidisciplinary Journal*, 26(4): 576–590. MR3999542. doi: https://doi.org/10.1080/10705511.2018.1558060. 185, 186, 197

Schiavon, L., Canale, A., and Dunson, D. B. (2022). "Generalized infinite factorization models." *Biometrika*, 109(3): 817–835. MR4472850. doi: https://doi.org/10.1093/biomet/asab056. 199

Schwarz, G. (1978). “Estimating the Dimension of a Model.” *The Annals of Statistics*, 6(2): 461–464. MR0468014. 182

Thurstone, L. L. (1947). *Multiple-Factor Analysis: A Development & Expansion of The Vectors of Mind*. University of Chicago Press. MR0021272. 183, 185

Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288. MR1379242. 185

West, M. (2003). “Bayesian factor regression models in the “large p, small n” paradigm.” *Bayesian Statistics*, 7: 733–742. MR2003537. 182, 185, 200

Woody, S., Carvalho, C. M., and Murray, J. S. (2021). “Model interpretation through lower-dimensional posterior summarization.” *Journal of Computational and Graphical Statistics*, 30(1): 144–161. MR4235972. doi: https://doi.org/10.1080/10618600.2020.1796684. 182, 184

Zhang, C.-H. et al. (2010). “Nearly unbiased variable selection under minimax concave penalty.” *The Annals of statistics*, 38(2): 894–942. MR2604701. doi: https://doi.org/10.1214/09-AOS729. 186

Acknowledgments

The authors are grateful to an Associate Editor and two anonymous Referees for all their comments, corrections, and suggestions which improved remarkably the paper.